BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37034
doi: 10.3897/biss.3.37034

OPEN ACCESS

Conference Abstract

‘The Last Mile’: The registry behind the identifier

Alex R Hardisty[‡], Larry Lannom[§], Dimitris Koureas[|,¶], Wouter Addink[|,¶], Claus Weiland[#]

‡ Cardiff University, Cardiff, United Kingdom
§ CNRI, Reston, VA., United States of America
| Naturalis Biodiversity Center, Leiden, Netherlands
¶ Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands
# Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Frankfurt, Germany

Corresponding author: Alex R Hardisty (hardistyar@cardiff.ac.uk)

Received: 10 Jun 2019 | Published: 13 Jun 2019

Citation: Hardisty A, Lannom L, Koureas D, Addink W, Weiland C (2019) ‘The Last Mile’: The registry behind the identifier. Biodiversity Information Science and Standards 3: e37034. https://doi.org/10.3897/biss.3.37034

Abstract

Preserved specimens in natural science collections have lifespans of many decades and often, several hundreds of years. Specimens must be unambiguously identifiable and traceable in the face of changes in physical location, changes in organisation of the collection to which they belong, and changes in classification. When digitizing museum collections, a clear link must be maintained between the physical specimen itself and the information digitally representing that specimen in cyberspace. The idea of a Natural Science Identifier (NSId) as a neutral, unique, universal and stable long-term persistent identifier (PID) of a ‘Digital Specimen’ is central to museums’ ambitions for widening access. An NSId allows easy identification and referencing of specific Digital Specimens, regardless of type, location, owner or user. It provides a digital doorway to physical specimens through which services for arranging loans and visits can be accessed, as well as opening the door to innovative services for manipulating specimens’ information directly; for work reliant upon discovery of related third-party information; and for demanding 3D modelling and visualization of specimens. Because the work takes place within e-Infrastructures/Cyberspace, new possibilities for analysing hundreds of thousands of specimens simultaneously are opened by exploiting large-scale cloud computing capacity and deep mining/machine learning, for example.

There are several established identifier mechanisms that could be used as a basis for NSId, but some variant of Handles is most appropriate over the very long-term because of

© Hardisty A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

their neutrality, resistance to change and sustainability. Adopted uses of the Handle system include identification of journal articles and datasets in education and research (using Digital Object Identifiers); film and television programme assets in the entertainment sector; financial derivatives; and for international shipping and construction.

Aside from being stable and sustained over time, an essential requirement of a global PID mechanism is independence from the museums/institutions assigning identifiers. NSIds are opaque insofar as no information can or should be inferred solely by inspecting the identifier. Stakeholders change, collections move, and organisations evolve, merge or disappear. Even designations and descriptions of specimens and collections can change. Information should only be revealed when the identifier is resolved via a neutral index.

One can debate the most appropriate instantiation of the Handle system but this is not useful. Relevance, ease of use and added-value of the supporting 'NSId Registry' (NSIdR) – the index of the different kinds of natural science object and their relations – are the decisive factors. This can be seen from the example of the Entertainment Identifier Registry (EIDR) founded by the major motion picture studios to create a reliable way to identify and track film and TV content distribution. Focus on the object model, promotional branding and value perception in the target user segment are the critical factors for success. Providing such a registry, seamlessly coupled to work practices and language of the professionals addresses the last mile challenge (Koureas et al. 2016).

From specimens, class characteristics, storage containers and collections, to specific identifications, images, naming, literature references and more, the NSIdR's triple-hierarchy object model, rooted in OBO Foundry's Biological Collections Ontology, is the key to persistently identifying, relating and indexing the entire range of collection objects of interest to scientists and others working in the bio and geo realms. The NSIdR 'knowledge graph', interoperable with other identifier schemes, supports novel first- and third-party value-add services such as arranging loans and visits, curation and annotation, and machine-learning for relationship discovery and pattern exploration.

Keywords

persistent identifier, registry, Digital Object Architecture, handle

Presenting author

Alex R Hardisty

Presented at

Biodiversity_Next 2019

Funding program

Horizon 2020 Framework Programme of the European Union, H2020-INFRADEV-2016-2017 Grant Agreement No. 777483

Grant title

ICEDIG - Innovation and Consolidation for Large-Scale Digitisation of Natural Heritage

References

- Koureas D, Arvanitidis C, Belbin L, Berendsohn W, Damgaard C, Groom Q, Güntsch A, Hagedorn G, Hardisty A, Hobern D, Marcer A, Mietchen D, Morse D, Obst M, Penev L, Pettersson L, Sierra S, Smith V, Vos R (2016) Community engagement: The 'last mile' challenge for European research e-infrastructures. Research Ideas and Outcomes 2: e9933. https://doi.org/10.3897/rio.2.e9933